Understanding cultural impact in music and politics, using Wikimedia as a lens.

Brett Buttliere
University of Warsaw

Abstract

The goal of the project is to understand cultural impact music and politics by examining the words used to describe the (work of) public individuals, in relation to how often the profile is viewed. We examine especially the prevalence and mixture of positive and negative words, as indicated using Hu and Liu's sentiment dictionary from tidytext. Our team has run this analysis within the context of science and found that both positive and negative words in a profile are related to average monthly view counts. This work was positively reviewed at *Scientometrics*, and we expect it to be published soon.

The goal of this proposal and the bulk of its funds are to hire two ph.d. students, at 50% time, to use these same analysis scripts, replicating and extending the analysis from science into music and politics. In the end we expect to get at least one paper out of each work package (1 about music 1 about politics) and one more overarching study describing all three analyses in a top journal. This high quality use of Wikimedia data hopefully encourages others to use it. Our work more generally brings in researchers and engages science with Wikimedia.

Introduction

Our main goals are to understand how cultural impact works and normalize the usage of Wikimedia data for answering historical and cultural questions.

The proposal replicates and extends an existing project which analyzed approximately 100,000 Scientists that have a profile on Wikipedia, finding that both positive and negative sentiment, but especially a mix of the two, are predictive of average monthly pageviews (Buttliere, et al., 2023). Two work packages correspond to two students at 50% tim working to replicate and extend these findings:.

WP1: Analysis of Politicians on Wikipedia: Emphasizing improving sentiment analysis.
WP2: Analysis of Musicians on Wikipedia: Emphasizing improving visualization of data.

The goal will be to produce at least two papers of high interest and evidential value, as we believe the initial study of science, being published at *Scientometrics,* will be. Such a study not only mentions Wikimedia and these notable people in the same sentences in (e.g., news coverage), it positions Wikimedia as an important source of data and for testing many questions about e.g., communication, history, and in general questions about what is salient in the public consciousness, when. Demonstrating this will bring new researchers to the field.

Date: June 1, 2024 - May 30, 2025

Related work

There exist many studies working to understand impact and cultural change in e.g., Science (Kuhn, 1957), but few using Wikimedia, and even fewer examining the role of controversy and positive and negative sentiment. Our initial project sought to test Kuhn's ideas at Wiki Scale, and the idea now is to extend these ideasfrom science to politics and musical culture. The PI (Dr Buttiere's) work since at least 2013 has been about science communications online (Buttliere & Buder 2017; Buder, et al., 2023), and more recently about how Wikimedia can help solve the problems of science, e.g., making it open.

Methods

The study uses the words in a Wikimedia biography of a public person (scientist, politician, musician), and how often that profile is viewed on average per month. One project will start with the root category 'musicians by nationality' and the other 'politicians by nationality' as sampling frames, which should sample ~100,000+ profiles.

In the original study we used the well regarded Hu and Liu (2004) sentiment analysis dictionary, which is implemented in the tidytext() package (Silge & Robinson, 2016). We also used the the VADER Negation dictionary (Hutto & Gilbert, 2014).

Descriptive and correlation analyses are used to understand relationships between these indicators and the metric of attention.

The idea is to extend this basis.

Expected output

We expect to have 1 presentation and 1 paper per student. Our team is based in Poland and Wikimania is in Krakow this year. We intend to be there, though we doubt the project will have results at that point..

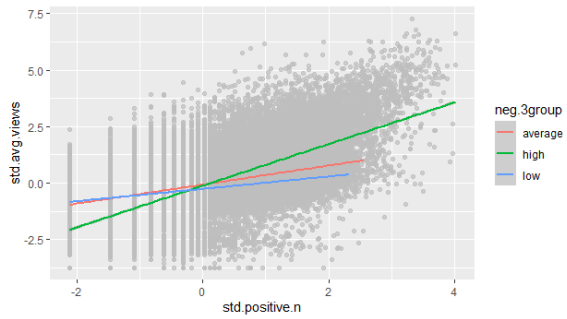

Figure 1: Positive sentiment (x), negative sentiment (color), and average monthly views (y) across 97,909 profiles of scientists.

Risks

We consider the study as of relatively low risk, given that all of the data is already openly available and it has already been run successfully once, also through peer review, without any major problems.

Community impact plan

The goal is to create relevant demonstrations of Wikimedia usage. When these studies are released we will be pushing them not only in social media but also through our editorial networks, and we hope that they will get picked up to some extent by the media.

Evaluation

Within 1 year PI will regard it as a success if we have presented it one time per student and submitted a paper each for publication. These papers are not likely to be published in this year. The papers should be examined for scientific value. Submitting another grant to a non Wikimedia grantor would be another indicator of success.

Budget

- 12,000 to Student 1- Leading WP1, 1,000 per month 50% contract (it is a good student wage, rent ~ 300$).
- 12,000 to Student 2- Leading WP2, 1,000 per month 50% contract (it is a good student wage, rent ~ 300$).
- 12,000 to PI - 1 day per week, administering project, ~20% work.
- Institutional overhead $7,500
- Conference and travel expenses $4,500 - 1,500 to go to one conference each.

Total, 48,000

Prior contributions

Brett Buttliere has been actively researching in the area of meta science since 2014. He has published multiple papers on digital infrastructure in science (Buttliere, 2014), metrics and altmetrics in science (Buttliere & Buder, 2017), meta-data (Buttliere, 2021), and impact and psychology in science (Buttliere, et al., 2023).

References

Buttliere, B., & Wicherts, J. (2013). What next for scientific communication? A large scale survey of psychologists on problems and potential solutions. *Master's Thesis*

Buttliere, B. T. (2014). Using science and psychology to improve the dissemination and evaluation of scientific work. *Frontiers in computational neuroscience, 8,* 82.

Buttliere, B., & Buder, J. (2017). Personalizing papers using Altmetrics: Comparing paper 'Quality'or 'Impact'to person 'Intelligence 'or 'Personality'. *Scientometrics, 111*(1), 219-239.

Buder, J., Zimmermann, A., Buttliere, B., Rabl, L., Vogel, M., & Huff, M. (2023). Online interaction turns the congeniality bias into an uncongeniality bias. Psychological Science, 34(10), 1055-1068.

Hu, X., Tang, J., Gao, H., & Liu, H. (2013, May). Unsupervised sentiment analysis with emotional signals. In Proceedings of the 22nd international conference on World Wide Web (pp. 607-618).

Kuhn, T. S. (2012). *The Structure of Scientific Revolutions.* University of Chicago press.

Silge, J., & Robinson, D. (2016). tidytext: Text mining and analysis using tidy data principles in R. *Journal of Open Source Software, 1*(3), 37.